

AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences

Year: 2023 | Citations: 1110 | Authors: M. Váradi, Damian Berton, Paulyna Magana, Urmila Paramval, Ivanna Pidruchna

Abstract

Abstract The AlphaFold Database Protein Structure Database (AlphaFold DB, <https://alphafold.ebi.ac.uk>) has significantly impacted structural biology by amassing over 214 million predicted protein structures, expanding from the initial 300k structures released in 2021. Enabled by the groundbreaking AlphaFold2 artificial intelligence (AI) system, the predictions archived in AlphaFold DB have been integrated into primary data resources such as PDB, UniProt, Ensembl, InterPro and MobiDB. Our manuscript details subsequent enhancements in data archiving, covering successive releases encompassing model organisms, global health proteomes, Swiss-Prot integration, and a host of curated protein datasets. We detail the data access mechanisms of AlphaFold DB, from direct file access via FTP to advanced queries using Google Cloud Public Datasets and the programmatic access endpoints of the database. We also discuss the improvements and services added since its initial release, including enhancements to the Predicted Aligned Error viewer, customisation options for the 3D viewer, and improvements in the search engine of AlphaFold DB.